HKL2MAP 0.3a-beta

User Guide

Thomas R. Schneider

European Molecular Biology Laboratory - Hamburg Outstation

webapps.embl-hamburg.de/hkl2map

Contact: thomas.schneider.at.embl-hamburg.de

September 21, 2011

Contents

1 Intro **4**

1.1 HKL2MAP 4

1.2 Organization of this guide 4

2 Installation **5**

2.1 Current versions 5

2.2 HKL2MAP tarball 5

2.3 Other programs required 5

2.4 Installation on Linux and MAC 5

2.5 Windows 6

3 General Notes **7**

3.1 Starting HKL2MAP 7

3.2 Definition of Project 7

3.3 SHELXC 9

3.3.1 What type of experiment should I choose ? 9

3.3.2 Loading data 9

3.3.3 Graphics: $< I/sig(I) >$ vs. resolution 9

3.3.4 Graphics: Completeness vs. resolution 9

3.3.5 Graphics: $< d"/sig(d") >$ vs. resolution 9

3.3.6 Graphics: $CC(\Delta F_i, \Delta F_j)$ vs. resolution 10

3.4 SHELXD 10

3.4.1 The Number of sites to request 10

3.4.2 The resolution limit for SHELXD 11

3.4.3 Special position flag for SHELXD 11

3.4.4 Number of trials 11

3.4.5 SHELXD's figures of merit 11

3.4.6 Minimum distance between SHELXD sites 11

3.5 SHELXE 12

3.5.1 Invert substructure 12

3.5.2 Estimation of the solvent content 12

3.5.3 Auto-tracing 12

3.5.4 Graphical Display with Coot 12

3.5.5 SHELXE-OUTPUT: CC(Eobs,Ecalc) 12

3.5.6 CC against native 12
3.6 How to continue 12
3.7 Background literature 14

4 MAD phasing of TE-II **15**
4.1 Diffraction Data 15
4.2 Running SHELXC 15
4.3 Running SHELXD 17
4.4 Running SHELXE 18
4.5 Displaying Electron Density 19

5 Sulphur-SAD phasing of Zn-free insulin 21
5.1 Diffraction Data 21
5.2 Starting HKL2MAP and Entering the Project Name 21
5.3 Running SHELXC 21
5.4 Running SHELXD 22
5.5 Running SHELXE 24

6 Various 26
6.1 Tips and Tricks 26
6.1.1 How can I import data from mtz-format? 26
6.2 Known issues 26

7 Acknowledgements 26

1 Intro

1.1 HKL2MAP

HKL2MAP is a graphical user-interface for macromolecular phasing. The current version connects several programs from the SHELX-suite to guide the user from analyzing scaled diffraction data (SHELXC), via substructure solution (SHELXD) and phasing (SHELXE) to displaying an electron density map (COOT). HKL2MAP can handle SAD, SIR, SIRAS or MAD data.

HKL2MAP can be obtained from thomas.schneider.at.embl-hamburg.de.

The author does not take any responsibility for the correct function of HKL2MAP or any of the programs called by it. Reports on faulty, unexpected or destructive behaviour to the above email-address will be appreciated.

Information about the SHELX-suite of programs can be found at

`shelx.uni-ac.gwdg.de/SHELX`

Information about COOT can be found at:

`http://lmb.bioch.ox.ac.uk/coot/`

Please cite the following references when you report the use of HKL2MAP:

HKL2MAP**:** Pape T, Schneider TR (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. J. Appl. Cryst. 37:843-844.

SHELXC**:** Sheldrick GM (2003). Gö ttingen University.

SHELXD**:** Schneider TR, Sheldrick GM (2002). Substructure solution with SHELXD, Acta Cryst. D58:1772-1779.

SHELXE**:** Sheldrick GM (2002). Macromolecular phasing with SHELXE, Z. Kristallogr. 217:644-650.

COOT**:** Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot, Acta Cryst. D66:486-501

1.2 Organization of this guide

Installation on linux, MAC, Windows and how to start the program are described in Section 2.

General guidelines for setting the parameters accessible by via the HKL2MAP-interface are given in section 3 along with references explaining the underlying methodology.

The remaining sections describe different phasing scenarios:

- MAD phasing
- S-SAD Elastase - single sulphurs
- S-SAD Ins - split super-sulphurs
- I-SAD Elastats - soak
- Pt-MIR Lysozyme

2 Installation

2.1 Current versions

The applications in this guide have been run using Coot version 0.6.1 and SHELXC/D/E as available on the SHELX-server on 19/08/2011.

2.2 HKL2MAP tarball

The installation will normally be done starting from a tar-ball `hkl2map -0.3a-beta-dist` obtained either from the HKL2MAP-website after registering or from Thomas via email.

2.3 Other programs required

To run HKL2MAP you need to have the following programs in your PATH:

- `shelxc, shelxd, shelxe`
- `wish`. This is the interface to the Tcl/Tk toolkit, the programming language in which HKL2MAPis actually written. Most computers have wish in the PATH. Especially if your are using the CCP4i-interface, you will have `wish` in your PATH automatically. In case that `wish` is not available, ask your friendly IT guys for help.
- `coot`.

To test whether for example the executable `shelxc` is in your PATH, type:

```
% which shelxc
```

from the command-line.

2.4 Installation on Linux and MAC

Unpack the tarball in a temporary location:

```
% tar -xvzf hkl2map-0.3a-dist.tgz
```

Rename and copy the HKL2MAP-executable to a convenient location (here: `/Users/trs/bin`.

```
% cp hkl2map-0.3a-dist.tcl /Users/trs/bin/hkl2map.tcl
```

To make HKL2MAP run on your particular operating system, you may need to remove the first two lines of the script:

```
#!/usr/bin/env wish
# The following for non-MAC
#!/usr/bin/wish
```

such that

```
#!/usr/bin/wish
```

becomes the first line of the script. If HKL2MAP still does not work, ask your IT-guys whether `wish` is available on your system and if so where it lives.

To add the chosen location to your PATH, add the following line to your `.cshrc`-file:

```
setenv PATH /Users/trs/bin:${PATH}
```

On the MAC, you need to make sure that the runtime-libraries for the SHELX-programs can be found. For this, add the following line to your `.cshrc`-file (here the libraries reside in `/Users/trs/bin`):

```
setenv DYLD_LIBRARY_PATH /Users/trs/bin
```

2.5 Windows

I do not have a windows mashine in operation myself. I am grateful for any hints on this.

3 General Notes

3.1 Starting HKL2MAP

Once HKL2MAP is available in the path of executables, it can be started from the command line:

```
% hkl2map &
```

The first thing that HKL2MAP will do is to check what type of executables are available and print the respective information to the terminal window:

```
trs-laptop2:~/A-shelxi/exercise] tschneider% ./hkl2map

 ************************************************
 * HKL2MAP - Version 0.2j.0                     *
 ************************************************

 OS: Darwin 10.8.0 i386 i386
 Tcl/Tk: Version 8.5
 Windowing system: aqua

 GetExecutablePath for 'shelxc'.
 ==> Path is '/Users/tschneider/A-shelxi/NewShelx-110901/intelmac/shelxc'.
 ==> Version is '2011/1'

 GetExecutablePath for 'shelxd'.
 ==> Path is '/Users/tschneider/A-shelxi/NewShelx-110901/intelmac/shelxd'.
 ==> Version is 'SHELXD-2011/5'
 ==> This is an MP-enabled version of SHELXD running on   2  threads

 GetExecutablePath for 'shelxe'.
 ==> Path is '/Users/tschneider/A-shelxi/NewShelx-110901/intelmac/shelxe'.
 ==> Version is '2011/1'

 GetExecutablePath for 'coot'.
 ==> Path is '/usr/local/xtal/coot64/bin/coot'.
 ==> Version is '0.6.1'

 GetExecutablePath for 'xterm'.
 ==> Path is '/usr/X11/bin/xterm'.
 ==> Version is ''
```

Please have a look and check whether everything makes sense.

After closing the Welcome-Window, the Main-Window of HKL2MAP becomes visible (Figure 1). At the top of the Main-Window, a Menu-Bar and an input line for the project name are located. For the different programs called by HKL2MAP, the interface provides Panels that can be openend and closed by clicking the respective boxes. The bottom part of the Main-Window consists of a Status-Window containing the output generated by the called programs and a line displaying brief context-sensitive help-messages.

3.2 Definition of Project

In the context of HKL2MAP, we consider an attempt to phase a structure with a certain set of parameters as a 'Project'. The project name specified in HKL2MAP will be used as a prefix for all files generated by HKL2MAP or the SHELX-programs called by it.

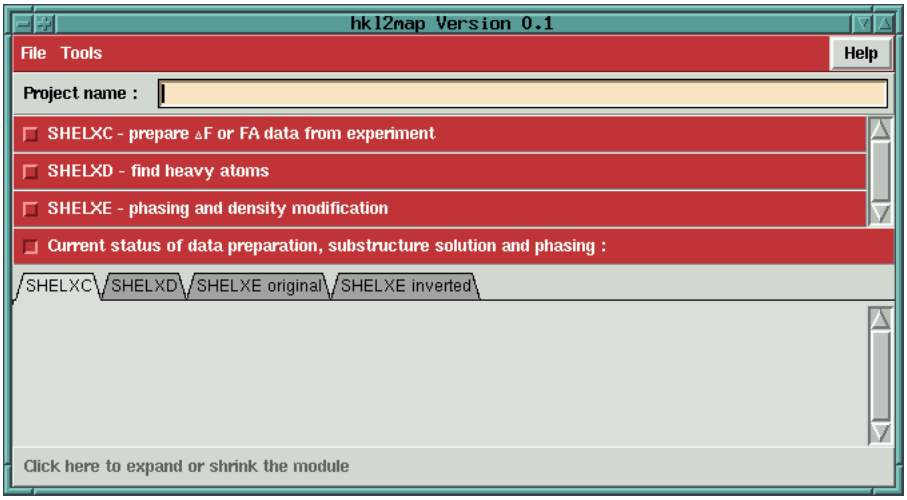
hkl2map Version 0.1
File Tools
Help
Project name :
SHELXC - prepare ΔF or FA data from experiment
SHELXD - find heavy atoms
SHELXE - phasing and density modification
Current status of data preparation, substructure solution and phasing :
SHELXC
SHELXD
SHELXE original
SHELXE inverted
Click here to expand or shrink the module

Figure 1: HKL2MAP Main-window after starting the program.

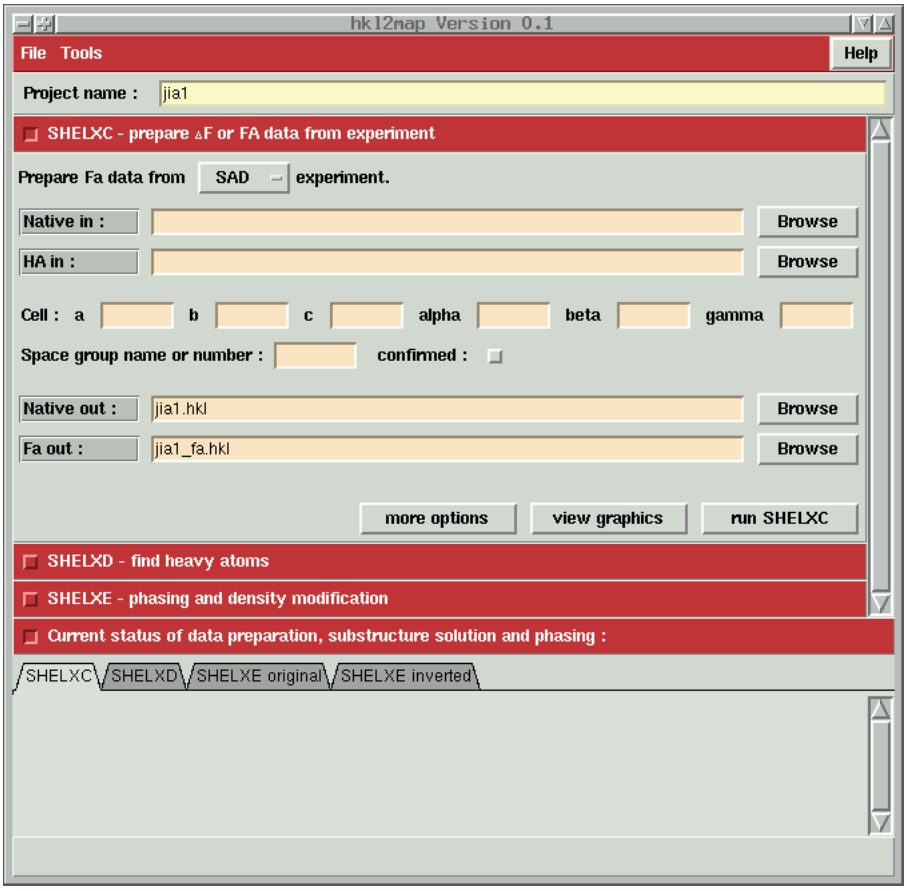

Figure 2: HKL2MAP Main-window after entering the project name.

Once the project name has been entered, all input-fields of the interface concerning input- and output-filenames are loaded with default values derived from the project name, and the SHELXC-Panel is opened (Figure 2).

3.3 SHELXC

SHELXC is a program to analyse diffraction data originating from SAD/MAD/SIRAS/MIRAS experiments. It prints some useful statistics to access the quality of the data and extracts substructure structure factors (DeltaF's or FA's) that represent the substructure of the heavy atoms or the anomalous scatterers only. These substructure structure factors can then be used by SHELXD to find the substructure atoms.

SHELXC also estimates an angle alpha - the angle between the complex structure factors of the protein atoms (FP) and the heavy atoms (FH) or the anomalous scatterers (FA). This angle will later be used by SHELXE to start the calculation of phases for the protein.

Generally, it is recommended to read scaled but unmerged data as the statistics generated from these data will be more reliable.

3.3.1 What type of experiment should I choose ?

The first time around, you will probably choose the type of experiment that you intended to do in the first place. However, if it turns out that the data are not of the expected quality, one can divert to other strategies. A typical example would be a MAD-experiment where the crystal suffered from radiation damage and only the data set collected first can be used in a SAD-approach.

3.3.2 Loading data

On input, SHELXC can read diffraction data in SHELX .hkl or in SCALEPACK .sca formats. Depending on the type of experiment chosen, the appropriate fields will be displayed for defining filenames.

3.3.3 Graphics: $< I/sig(I) >$ vs. resolution

This plot shows the signal-to-noise for the measured diffraction intensities (different data sets in different colors) versus resolution. As diffracted intensities are less at higher resolution than at low resolution, they become more difficult to measure, so the plot signal-to-noise ($< I/sig(I) >$) for the measurements should generally go down with increasing resolution.

3.3.4 Graphics: Completeness vs. resolution

For a good dataset, the completeness should be above 95% for all resolution bins. Sometimes the completeness is lower at low resolution - this can be due to overloaded reflections being rejected or a too small rotation range. For the substructure solution, incomplete data are sometimes acceptable, for phasing, it is very important to have complete data.

3.3.5 Graphics: $< d" / sig(d") >$ vs. resolution

This plot shows the signal-to-noise for the anomalous differences. As the anomalous differences are measured as small differences between large numbers (the differences in intensity between Friedel-mates) the signal-to-noise for the anomalous differences is usually not very high. This quantity also goes down with increasing resolution.

Anomalous data contain useful information as long as the d" calculated by SHELXC is higher than 0.8.

If data have been collected at several wavelengths, the $< d"/sig(d") >$ should be higher for the wavelengths with larger anomalous signal, e.g. PEAK>HREM>INFL>LREM for a typical Se-Met experiment.

3.3.6 Graphics: $CC(\Delta F_i, \Delta F_j)$ vs. resolution

To a first approximation, the anomalous differences measured at different wavelengths, should be related by the same constant factor for all reflections coming from the same crystal. Therefore, for ideal data, the correlation coefficient between anomalous differences measured at a wavelength i and at another wavelength j should be 1.0 or 100

In the real world, this is of course usually not the case. Nevertheless, the value of the correlation coefficient for data grouped in resolution bins is a very useful indicator of the accuracy of the measurements.

Generally, the correlation coefficient goes down for increasing resolution. A correlation coefficient of less than 30% is a good criterion for truncating the data at the respective resolution.

Beware that the correlation coefficient can be artifically high, if both data sets are not statistically independent or suffer from the same systematic errors.

For SAD data, the correlation coefficient can not be calculated between anomalous differences collected at different wavelengths. Instead, it is calculated by comparing different measurements of the same anomalous differences - note that this can only be done with unmerged data.

3.4 SHELXD

SHELXD is a program for the 'ab initio' solution of crystal structures. It can be used for solving entire structures from the diffraction intentisities alone or for solving substructures of heavy atoms and/or anomalous scatterers against the appropriate substructure structure factors. In the context of HKL2MAP, SHELXD is used to solve substructures.

3.4.1 The Number of sites to request

This number should be as close to the reality as possible.

For Se-Met derivatised crystals, it is usually possible to have a rather accurate idea on how many sites one should expect. If Se-incorporation has been checked for the protein preparation, the uncertainties in the number of sites originate from not knowing the number of molecules in the assymetric unit and not knowing whether or not all Methionine sidechains are ordered. Most people consider Seleniums in N-terminal Methionines as disordered.

For halide soaks, it is usually advantageous to start with a relatively small number, as spurious sites will be included in the calculation of figures of merit and render these meaningless in extreme cases. For phasing sites with and occupancy of less than 20-30 % are probably meaningless.

For disulphide bridges, if resolution is higher than 2.1 Å, one can look for single atoms. With data worse than 2.4 Å, it is better to look for super-atoms and then split these. Between 2.1 Åand 2.4 Å, try both options.

3.4.2 The resolution limit for SHELXD

The resolution limits are very important parameters for SHELXD. Direct methods are very sensitive to inaccurate data - a few badly measured reflections can make a substructure determination fail completely.

The low resolution cut-off should be chosen in a way, that overloaded data and reflections that were actually caught by the beamstop are excluded.

The high resolution cut-off should be chosen such that not too much noise is included. For multi-wavelength experiments a good high-resolution cutoff is where the correlationc coefficient between signed anomalous differences falls below 30For single-wavelength experiments the data can be truncated where the $< d" / sig(d") >$ falls below 1.4-1.5.

The number of reflections after application of E-value and/or resolution cut-offs should be at least 20/30 times larger than the number of expected sites.

3.4.3 Special position flag for SHELXD

For soaks, sites can end up on special positions, so the avoidance of special positions feature should be switched off.

3.4.4 Number of trials

Normally 100 trials is a good number.

For SAD data (as they are sometimes quite weak and noisy), it may be worth trying 1000 trials.

For very difficult cases, 50000 tries can be used. Large numbers of tries can be used especiallly when the MP-enabled version of SHELXD is used.

3.4.5 SHELXD's figures of merit

CC corresponds to an weighted correlation coefficient between E_{obs} and E_{calc}. The fact that it is WEIGHTED by 1/sigma(E) creates the danger, that its value can be strongly influenced by incorrect sigmas; i.e. a reflection with an extremely small sigma may dominate that whole evaluation of the CC.

PATFOM is not on any absolute scale. Also, for low symmetry space groups it may be deceivingly high and for high symmetry space groups (only few symmetry operations are checked in the Patterson minimum function, so noise may dominate), it may be deceivingly low (many symmetry operations are checked in the Patterson minimum function, one Zero is enought to result in a Zero PATFOM for a certain pair of atoms).

The occupancies obtained from SHELXD are renormalized such that the highest site has an occupancy of 1.0 (even if, in reality, the occupancy is only 0.5). This can lead to complications if even the highest occupancy site still has small occupancy, e.g. in halide soaks. After correcting for this overall scaling, the occupancies are probably correct within plus/minus 20

3.4.6 Minimum distance between SHELXD sites

Typical minium distances for Se: 3.5 A; S-S bonds: 1.7 A; Fe in Heme groups: 6 A;

3.5 SHELXE

SHELXE is a program to rapidly derive a phase set for data set based on substructure structure factors and a substructure consisting of heavy atoms and/or anomalous scatterers.

To find out what files SHELXE uses and what files it writes out, have a look at the top-part of the respective log-file

CPU-times for 'tl2' correspond to trs-laptop2. This is a 2.66 GHz Intel Core 2 Duo MAC BookPro.

3.5.1 Invert substructure

Will SHELXE deal properly with space groups that need changing when the structure is inverted ? Yes, of course ...

3.5.2 Estimation of the solvent content

The solvent content of the crystal under consideration is calculated by assuming a volume of 140 $Å^3$ per amino acid residue:

SC = # residues * 140 $Å^3$ / volume(a.u.),

whereby volume(a.u.) is the volume of the asymmetric unit.

3.5.3 Auto-tracing

If the CC after the first auto-tracing round is $>25\%$, the structure is usually solved.

3.5.4 Graphical Display with Coot

XXX To be written. Ctrl-Left-Mouse to translate the electron density.

3.5.5 SHELXE-OUTPUT: CC(Eobs,Ecalc)

The CC between Eobs and Ecalc printed for the heavy atoms substructure is calculated on all data (as apposed to the CC in SHELXD). If this value is $>15\%$, the situation is good. Note that this number depends a lot on resolution cut-offs.

3.5.6 CC against native

If the CC after the first auto-tracing round is $>25\%$, the structure is usually solved.

3.6 How to continue

To move the structure factors and phases into the CCP4-world, here is template/script:

```
#!/bin/csh -f
#
# phs2mtz
# -------
```

```
#
# C-shell script to convert phs-file produced by SHELXE
# into mtz-format. For running this script, you need to
# have the CCP4-programs 'f2mtz' and 'cad' in your path.
#
# Before running the script you have to adjust the unit
# cell parameters and the space group to the appropriate
# values.
#
# (C) Thomas R. Schneider, 22-Oct-2004
#
############################################################
echo ""
echo " Converting $1 ... "
echo ""
if (-e $1) then
  echo " The file $1 is available"
else
  echo " The file $1 is not available (did you use the extension ?)"
  echo ""
  exit
endif
# Check for programs
foreach v (f2mtz cad)
  echo -n " Checking for $v "
  if (`which $v | grep -ci "not found"`) then
    echo " - not found "
    echo " Exiting. "
    echo ""
    exit
  else
    echo " - found. OK. "
  endif
end
# Run f2mtz
echo ""
echo " Running f2mtz"
f2mtz hklin $1 hklout $1:r.mtz > t_$1:r_f2mtz.log <<END
cell 96.0 120.0 166.1 90.0 90.0 90.0
symmetry C2221
labout H K L FP FOM PHIB SIGFP
CTYPOUT H H H F W P Q
END
grep "Number of Reflections " t_$1:r_f2mtz.log
grep "Space Group" t_$1:r_f2mtz.log
echo -n " * Resolution Range: "
grep " A )" t_$1:r_f2mtz.log
# Run cad
echo ""
echo " Running cad "
cad hklin1 $1:r.mtz hklout $1:r.cad > t_$1:r_cad.log <<eof-cad
LABIN FILE 1 E1=FP E2=FOM E3=PHIB E4=SIGFP
LABOUT E1=FP E2=FOM E3=PHIB E4=SIGFP
eof-cad
# Report
echo ""
echo " -> Logs are in t_$1:r_*.log"
echo " -> Data are in $1:r.cad "
echo ""
```

3.7 Background literature

XXX To be written.

4 MAD phasing of TE-II

4.1 Diffraction Data

The data used for this tutorial originate from a four wavelengths MAD-dataset on a Se-Met derivative used to solve the structure of *Escherichia coli* thioesterase II [1] (TE-II). In addition to the four-wavelengths MAD-data, a native data set to 1.9 Å resolution was collected on the S-Met form of the protein. The data were kindly provided by Zbyzsek Dauter.

The structure of TE-II was determined in space group $C222_1$ with unit cell constants of: a = 96.0 Å, b = 120.0 Å, c = 166.1 Å, $\alpha = \beta = \gamma = 90.0°$. The data were processed with DENZO and SCALEPACK. The scaled and merged data were stored in SCALEPACK .sca-format.

code	λ [Å]	d_{min} [Å]	filename
hrm	0.9747	2.5	`jia_hrem.sca`
pk	0.9787	2.5	`jia_peak.sca`
ip	0.9793	2.5	`jia_infl.sca`
lrm	0.9801	2.5	`jia_lrem.sca`
nat	n/a	1.9	`jia.hkl`

Table 1: Overview of data sets used in this tutorial. `jia.hkl` is in SHELX-hkl4-format, i.e. contains $h, k, l, I, \sigma(I)$ in asci-format.

The refined structure of TE-II (pdb-code 1C8U) contains 2 x 4 Se-atoms in a total of 2 x 285 amino acids. The solvent content is ca. 66%. More information about the structure and its determination can be found in: Li et al. (2000) Nat.Stru.Biol. 7:555.

4.2 Running SHELXC

To run SHELXC, first the type of experiment has to be chosen (Figure 3). After the choice has been made, the SHELXC-panel changes and presents input fields appropriate for the respective type of experiment. For the present example, MAD has been selected and, in response, the interface generates input fields for filenames for files containing the diffraction data for a native and four MAD data sets. The filenames can be typed in or files can be selected using a browser that is actived by clicking on the Browser-Button.

As many 'accidents' happen due to misunderstanding between programs about space groups, HKL2MAP requires an explicit confirmation of unit cell constants and space group (Confirmed-Button) before SHELXC can actually be started from the interface.

Once SHELXC is started by clicking on the Run-SHELXC-Button, the interface is blocked for further input until the program has finished. The output of the SHELXC-run is directed to the Status-Panel at the bottom of the Main-Window.

After SHELXC has finished, a graphical version of the SHELXC-output can be activated by pressing the View-Graphics-Button. Clicking this button will open a graphics window allowing to display various plots. The type of plot to be displayed can be selected using the Display-Menu in the Graphics-Window.

The plots shown in the Graphics-Window can be written in postscript-format using the Make-Postscript-Option in the File-Menu of the Graphics-Window.

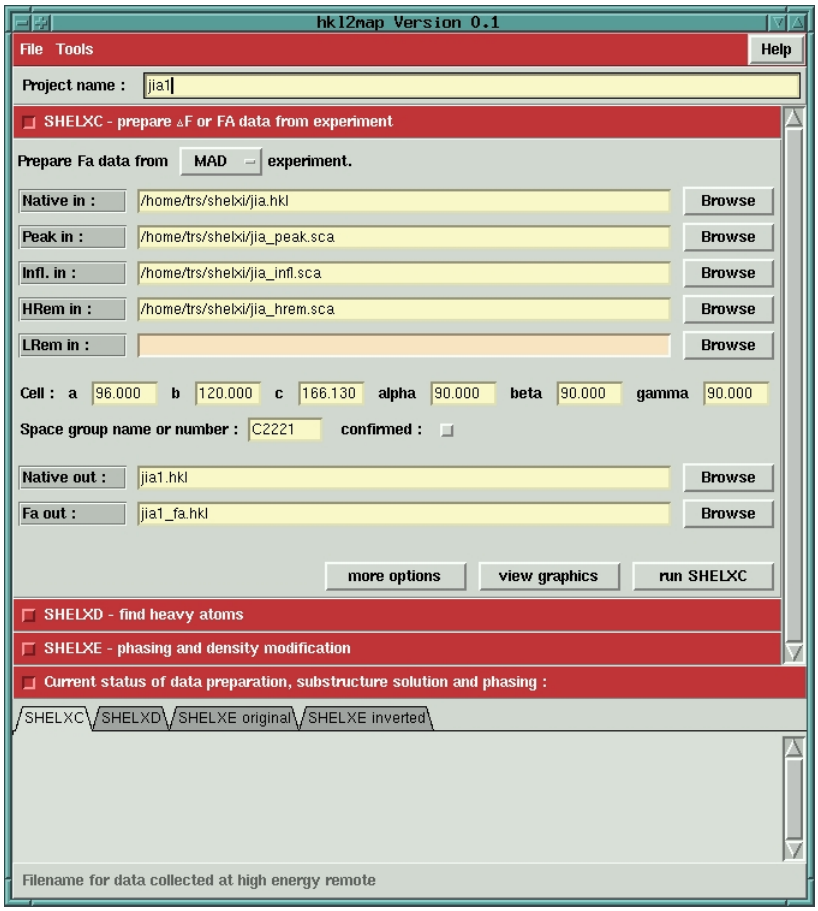
hkl2map Version 0.1
File Tools
Help
Project name : jia1
SHELXC - prepare ΔF or FA data from experiment
Prepare Fa data from MAD experiment.
Native in : /home/trs/shelxi/jia.hkl
Peak in : /home/trs/shelxi/jia_peak.sca
Infl. in : /home/trs/shelxi/jia_infl.sca
HRem in : /home/trs/shelxi/jia_hrem.sca
LRem in :
Browse
Cell : a 96.000 b 120.000 c 166.130 alpha 90.000 beta 90.000 gamma 90.000
Space group name or number : C2221 confirmed :
Native out : jia1.hkl
Fa out : jia1_fa.hkl
more options
view graphics
run SHELXC
SHELXD - find heavy atoms
SHELXE - phasing and density modification
Current status of data preparation, substructure solution and phasing :
SHELXC SHELXD SHELXE original SHELXE inverted
Filename for data collected at high energy remote

Figure 3: HKL2MAP Main-window after a MAD-experiment has been selected and names of files containing diffraction data have been entered.

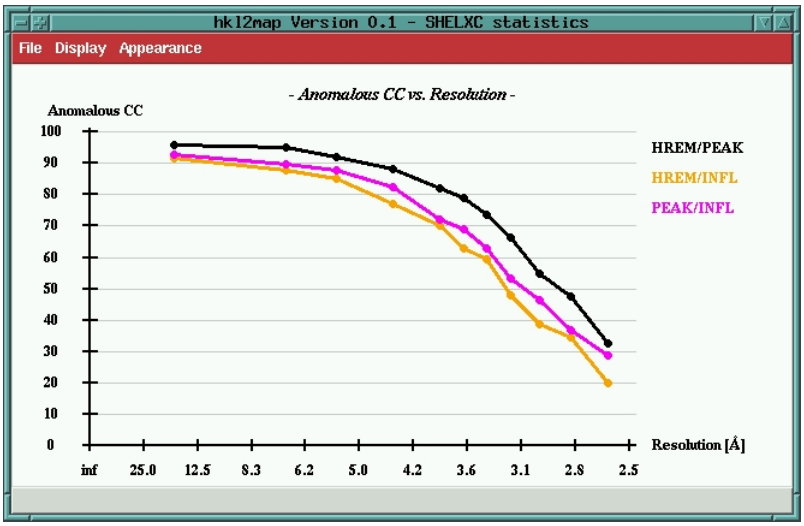

Figure 4: The Graphics-Window for the results of SHELXC after choosing the plot of 'Anomalous CC vs. Resolution'.

4.3 Running SHELXD

The name of the file containing substructure structures factors is automatically forwarded to the SHELXD part of the interface. Also, the interface tries to suggest a high resolution cut-off for the data used for finding the substructure.

The user has to supply the expected number of sites and the type of heavy atom in the crystal. In this case, 8 Se-sites are requested, and SHELXD is started by pressing the Run-SHELXD-Button.

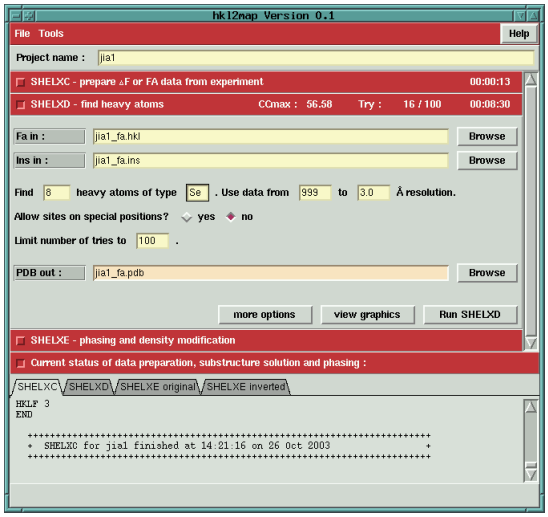

Figure 5: The Main-Window after entering the numbers necessary for starting a SHELXD-job.

The output from SHELXD is shown to the Status-Window at the Bottom of the Main-Window. Again, a graphical display of the results can be activated by pressing the View-Graphics-Button. As opposed to the SHELXC-related plots, the plots describing SHELXD are automatically updated to reflect the current status of the SHELXD-run and can be inspected at any time during the SHELXD-run.

One useful plot is the plot of peak-height vs. peak number for the current best solution (Figure 6). In favourable cases, this plot shows a clear drop indicating the last significant site of the substructure.

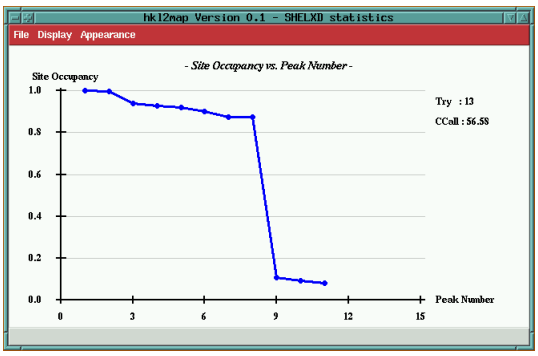

Figure 6: The Graphics-Window showing a plot of peak-heigt vs. peak-number. A clear drop in peak-height is seen between peak 8 and 9.

A SHELXD-run can be interrupted at any time using the Finish-SHELXD-Button.

4.4 Running SHELXE

The names of the files containing the native data, the substructure structure factors and the best sites found by SHELXD are automatically filled into the respective fields of the SHELXE-panel of the interface.

Before calculating phases with SHELXE, the fractional solvent content (for convenience a widget for automatic calculation of the solvent content can be activated by pressing the Estimate-Solvent-Content-Button) has to be entered. Also the user has to supply the information whether or not the native crystals contained the substructure atoms or not.

The interface allows to specify which enantiomorph of the substruture should be used for phasing. As an alternative to testing the enaniomorphs one after the other, the respective jobs can be run simultaneouly by choosing try both enantiomorphs. Dual-CPU computers are particularly convenient to run the jobs for the two enantiomorphs in parallel.

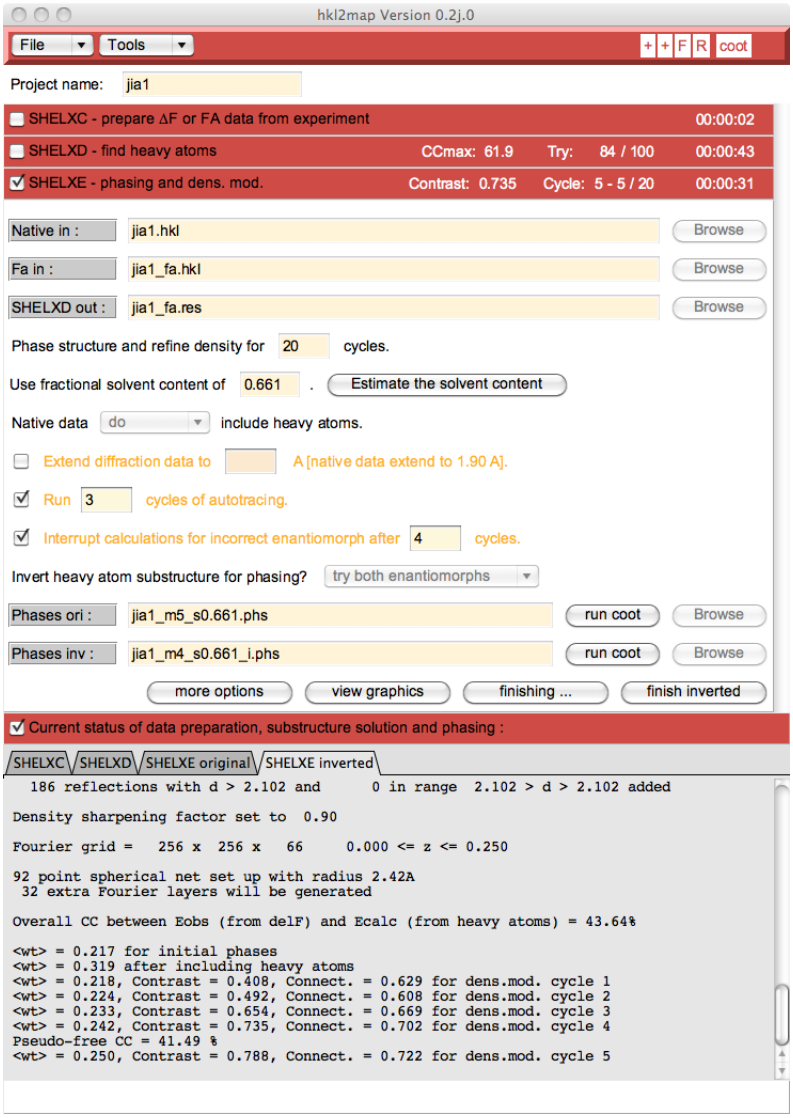
hkl2map Version 0.2j.0
File
Tools
+ + F R coot
Project name: jia1
SHELXC - prepare ΔF or FA data from experiment 00:00:02
SHELXD - find heavy atoms CCmax: 61.9 Try: 84 / 100 00:00:43
SHELXE - phasing and dens. mod. Contrast: 0.735 Cycle: 5 - 5 / 20 00:00:31
Native in : jia1.hkl Browse
Fa in : jia1_fa.hkl Browse
SHELXD out : jia1_fa.res Browse
Phase structure and refine density for 20 cycles.
Use fractional solvent content of 0.661 . Estimate the solvent content
Native data do include heavy atoms.
Extend diffraction data to A [native data extend to 1.90 A].
Run 3 cycles of autotracing.
Interrupt calculations for incorrect enantiomorph after 4 cycles.
Invert heavy atom substructure for phasing? try both enantiomorphs
Phases ori : jia1_m5_s0.661.phs run coot Browse
Phases inv : jia1_m4_s0.661_i.phs run coot Browse
more options view graphics finishing ... finish inverted
Current status of data preparation, substructure solution and phasing :
SHELXC SHELXD SHELXE original SHELXE inverted
186 reflections with d > 2.102 and 0 in range 2.102 > d > 2.102 added
Density sharpening factor set to 0.90
Fourier grid = 256 x 256 x 66 0.000 <= z <= 0.250
92 point spherical net set up with radius 2.42A
32 extra Fourier layers will be generated
Overall CC between Eobs (from delF) and Ecalc (from heavy atoms) = 43.64%
<wt> = 0.217 for initial phases
<wt> = 0.319 after including heavy atoms
<wt> = 0.218, Contrast = 0.408, Connect. = 0.629 for dens.mod. cycle 1
<wt> = 0.224, Contrast = 0.492, Connect. = 0.608 for dens.mod. cycle 2
<wt> = 0.233, Contrast = 0.654, Connect. = 0.669 for dens.mod. cycle 3
<wt> = 0.242, Contrast = 0.735, Connect. = 0.702 for dens.mod. cycle 4
Pseudo-free CC = 41.49 %
<wt> = 0.250, Contrast = 0.788, Connect. = 0.722 for dens.mod. cycle 5

Figure 7: The Main-Window after starting SHELXE.

The progress of the phasing calculations can be followed in the Status-Window or, more conveniently, by activating the graphical view of the SHELXE-output by pressing the View-Graphics-

Button. An important measure for phase quality is the 'Contrast' between the variance in the electron density in the protein region and in the solvent region for a given phase-set. Using the Graphics-Window activated via the View-Graphics-Button in the SHELXE-Panel, a continously updated plot of the value of the contrast vs. cycle number can be displayed and the decision which of the substructure-enantiomorphs results in the correct phases can be made while the phasing calculations are running.

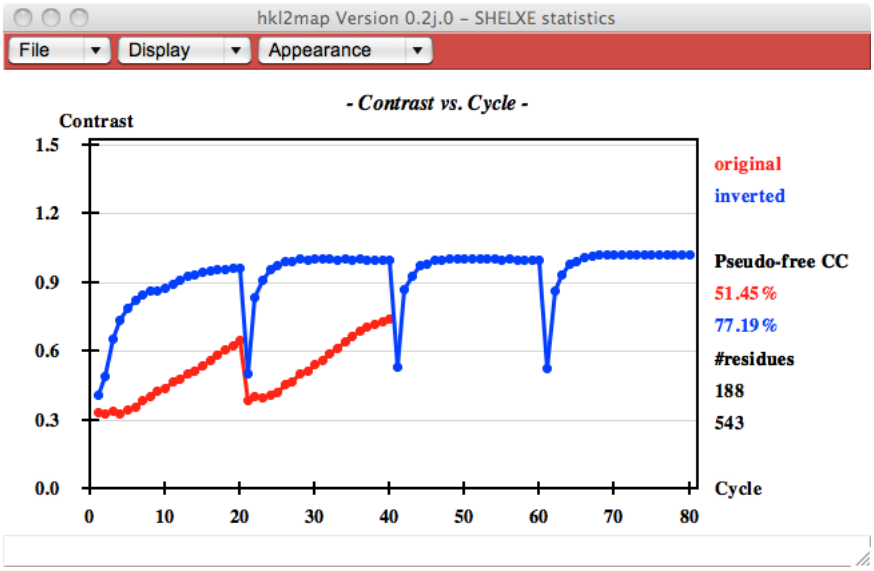

Figure 8: The plot of contrast vs. cycle number. The contrast-values for the original and the inverted substructure are shown in red and blue respectively. Here, the inverted substructure is the correct one. The phasing job for the incorrect substructure (red) is stopped halfway.

When a clear difference between contrasts for the two enantiomorphs is seen, the phasing job using the wrong enantiomorph can be interrupted by pressing the respective Finish-Original/Inverted-Button to free the computing resources.

4.5 Displaying Electron Density

The electron density maps corresponding to the phase sets produced by SHELXE can be inspected by pressing the run-coot-Button corresponding to the required substructure enantiomorph in the SHELXE-panel of HKL2MAP.

To display the result of the auto-tracing, load the pdb file `<proj>.pdb` (for the original substructure) or `<proj>_i.pdb` (for the inverted substructure) into COOT. For the case discussed here, the filenames would be `jia1.pdb` or `jia1_i.pdb`.

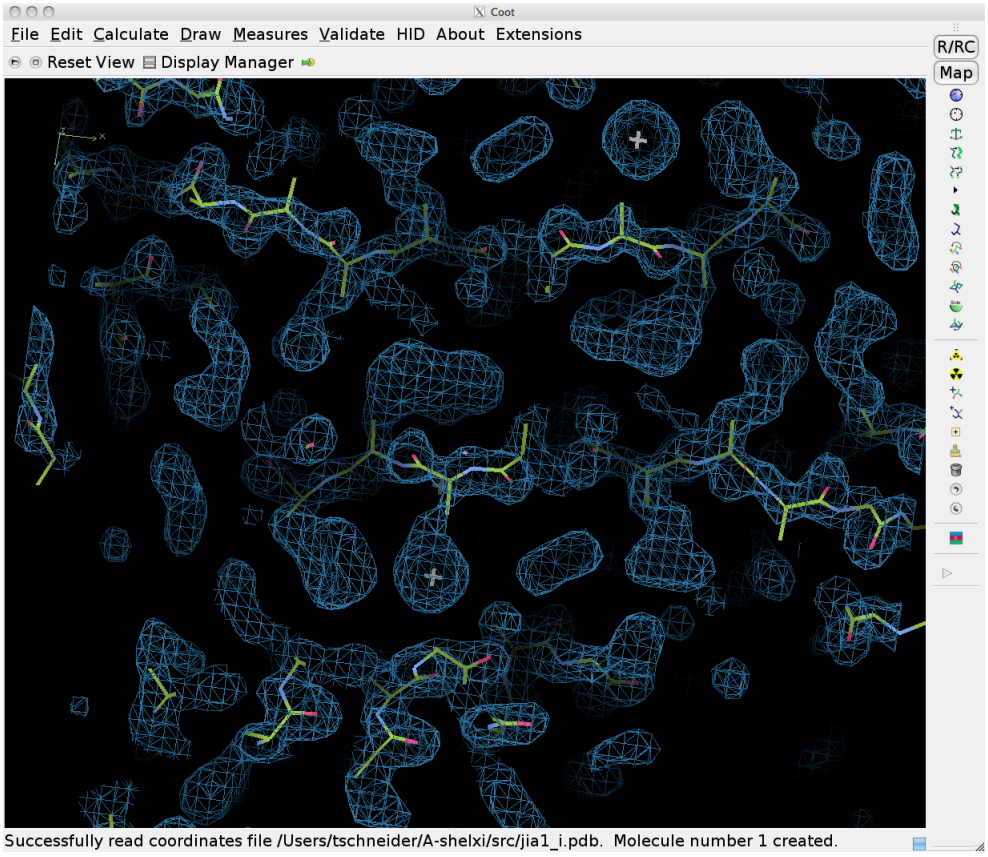
Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View Display Manager
R/RC
Map
Successfully read coordinates file /Users/tschneider/A-shelxi/src/jia1_i.pdb. Molecule number 1 created.

Figure 9: Electron density map for TE-II with auto-traced model superimposed as displayed by COOT.

5 Sulphur-SAD phasing of Zn-free insulin

5.1 Diffraction Data

The data used here originate from a test experiment on a crystal of Zn-free Insulin using the Cu-K_α-radiation emitted from the rotating anode in the Department of Experimental Oncology at the European Institute of Oncology in Milan, Italy. The crystal was grown following the instructions distributed by the Sheldrick laboratory for the BIOXHIT test crystal project.

The crystal belonged to space group I213 with unit cell parameters of $a = b = c = 78.3$ Å, $\alpha = \beta = \gamma = 90.0°$. 135 deg. of data were collected in increments of 1 deg. The data were processed with DENZO and SCALEPACK. The scaled, but unmerged data (keyword `no merge original index` in SCALEPACK) were stored in SCALEPACK .sca-format.

The crystal used contained one molecule of insulin (51 residues) per asymmetric unit. One molecule of insulin contains 3 disulphide bridges

5.2 Starting HKL2MAP and Entering the Project Name

HKL2MAP can be started from a UNIX-command line:

```
% hkl2map &
```

After closing the Welcome-Window, the Main-Window of HKL2MAP becomes visible (Figure 10). At the top of the Main-Window, a Menu-Bar and an input line for the project name are located. For the different programs called by HKL2MAP, the interface provides Panels that can be openend and closed by clicking the respective boxes. The bottom part of the Main-Window consists of a Status-Window containing the output generated by the called programs and a line displaying brief context-sensitive help-messages.

5.3 Running SHELXC

To run SHELXC, first the type of experiment has to be chosen. After the choice has been made, the SHELXC-panel changes and presents input fields appropriate for the respective type of experiment. For the present example, SAD has been selected and, in response, the interface generates input fields for filenames for files containing the diffraction data for a native and a SAD data set. The filenames can be typed in or files can be selected using a browser that is actived by clicking on the Browser-Button.

Please note that for this case the anomalous data and the native data were collected on the same crystal and are in fact in the same data set.

As many 'accidents' happen due to misunderstanding between programs about space groups, HKL2MAP requires an explicit confirmation of unit cell constants and space group (Confirmed-Button) before SHELXC can actually be started from the interface.

Once SHELXC is started by clicking on the Run-SHELXC-Button, the interface is blocked for further input until the program has finished. The output of the SHELXC-run is directed to the Status-Panel at the bottom of the Main-Window.

After SHELXC has finished, a graphical version of the SHELXC-output can be activated by pressing the View-Graphics-Button. Clicking this button will open a graphics window allowing to

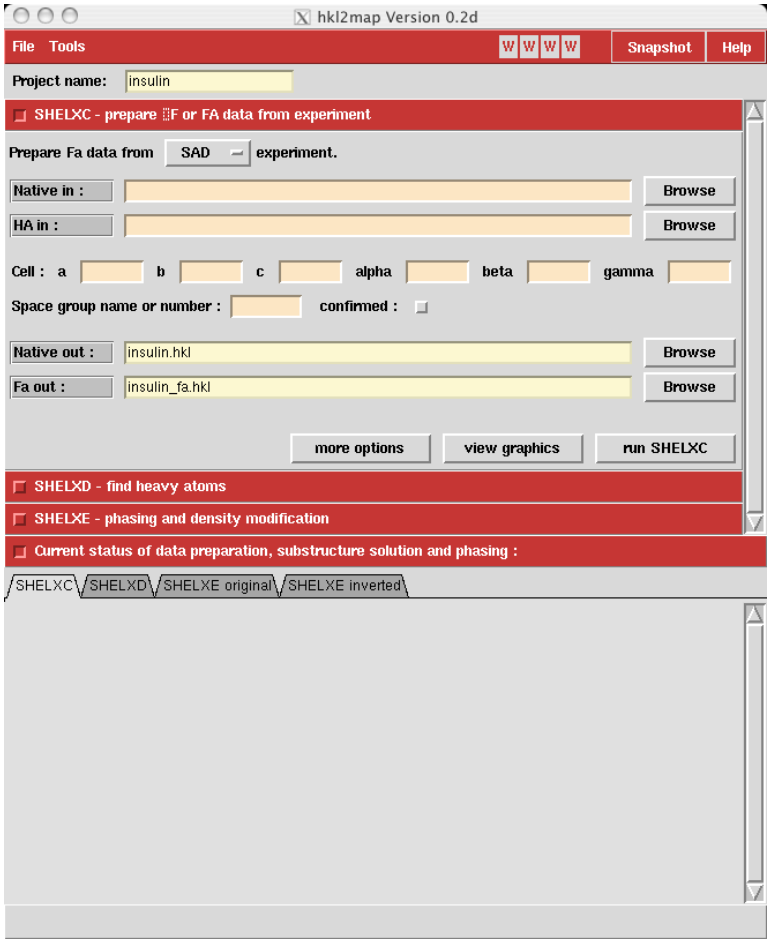

Figure 10: HKL2MAP Main-window after starting the program.

display various plots. The type of plot to be displayed can be selected using the Display-Menu in the Graphics-Window (Figure 12).

The signal-to-noise ratio for the anomalous signal as shown in the plot of d" vs. resolution (Fig. **??**.a) goes below 1.0 at ca. 2.2 Å, indicating that most of the data could actually be used for structure determination.

However, the anomalous correlation coefficients show a sharp drop around 2.7-2.9 Å, indicating that the data should better be truncated at this resolution. In addition the there is a sharp dip in the anomalous correlation coefficient between 3.3 and 4.0 Å, caused by diffraction measurements being affected by the presence of ice-rings.

The textual output of SHELXC is echoed in the bottom part of the Main-Window.

The plots shown in the Graphics-Window can be written in postscript-format using the Make-Postscript-Option in the File-Menu of the Graphics-Window.

5.4 Running SHELXD

Truncating the data to be used for substructure solution at 2.8 Å as suggested above will not allow to find the individual sulphur-atoms in disulfide bridges (ca. 2.0 Å distance). Therefore, SHELXD will first search for 3 'super-atoms' which then will be 'split' into two independent sulphur atoms. The 'splitting' is triggered by asking SHELXD to resolve 3 disulphide bridges in the part of the

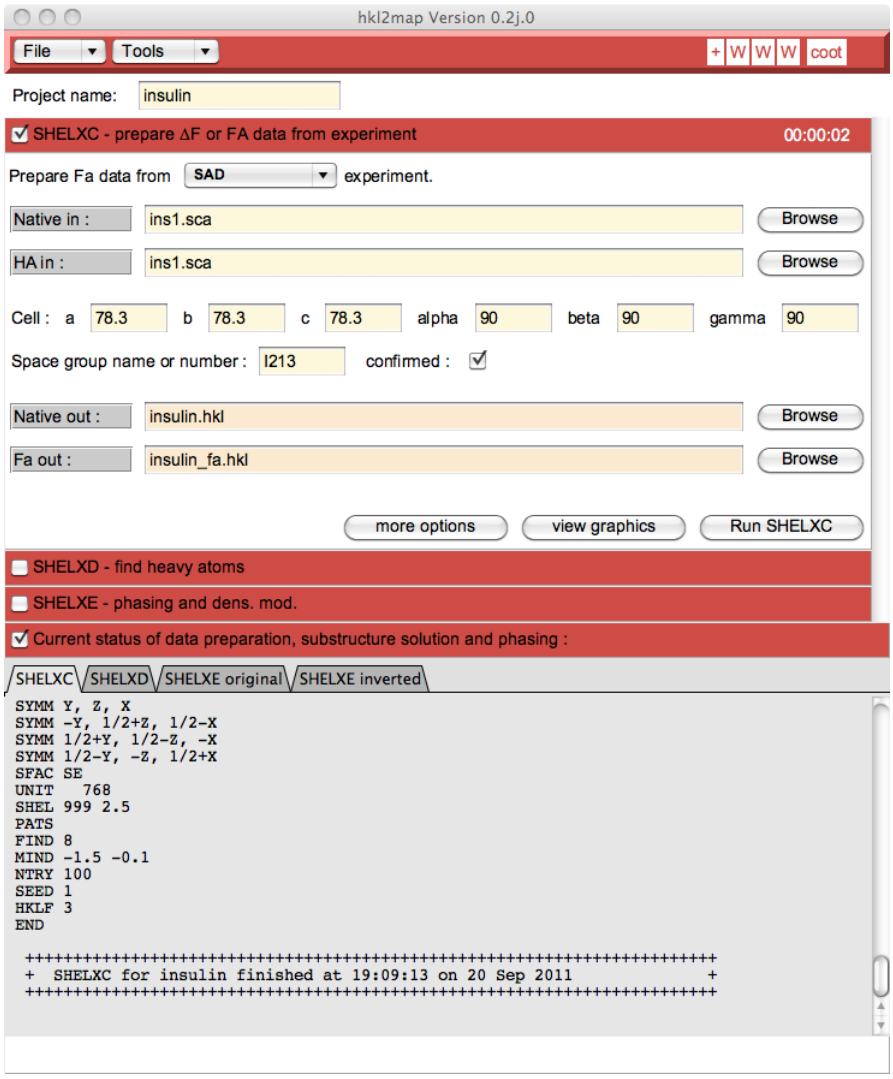

Figure 11: HKL2MAP Main-window after a SAD-experiment has been selected, the names of files containing diffraction data and the space group have been entered, and the SHELXC-job has been launched

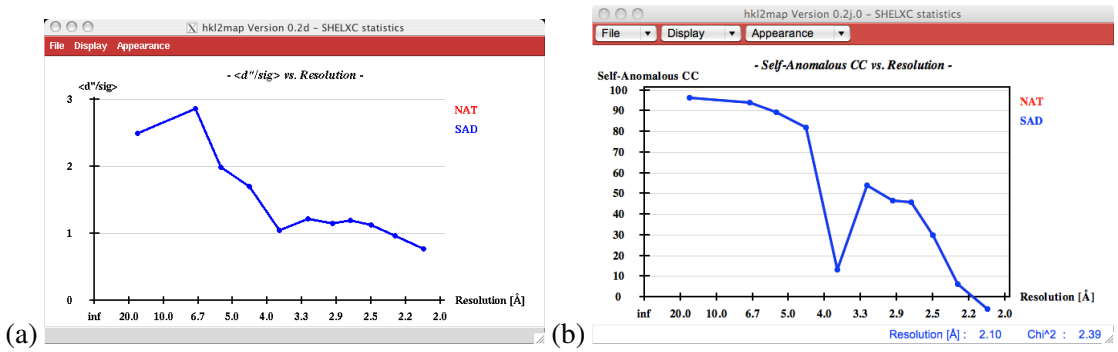

Figure 12: The Graphics-Window for the results of SHELXC after choosing (a) the plot of 'd" vs. Resolution' and (b) the plot of 'Anomalous CC vs. Resolution'.

SHELXD-panel which becomes visible when the More-Option-Button is pressed.

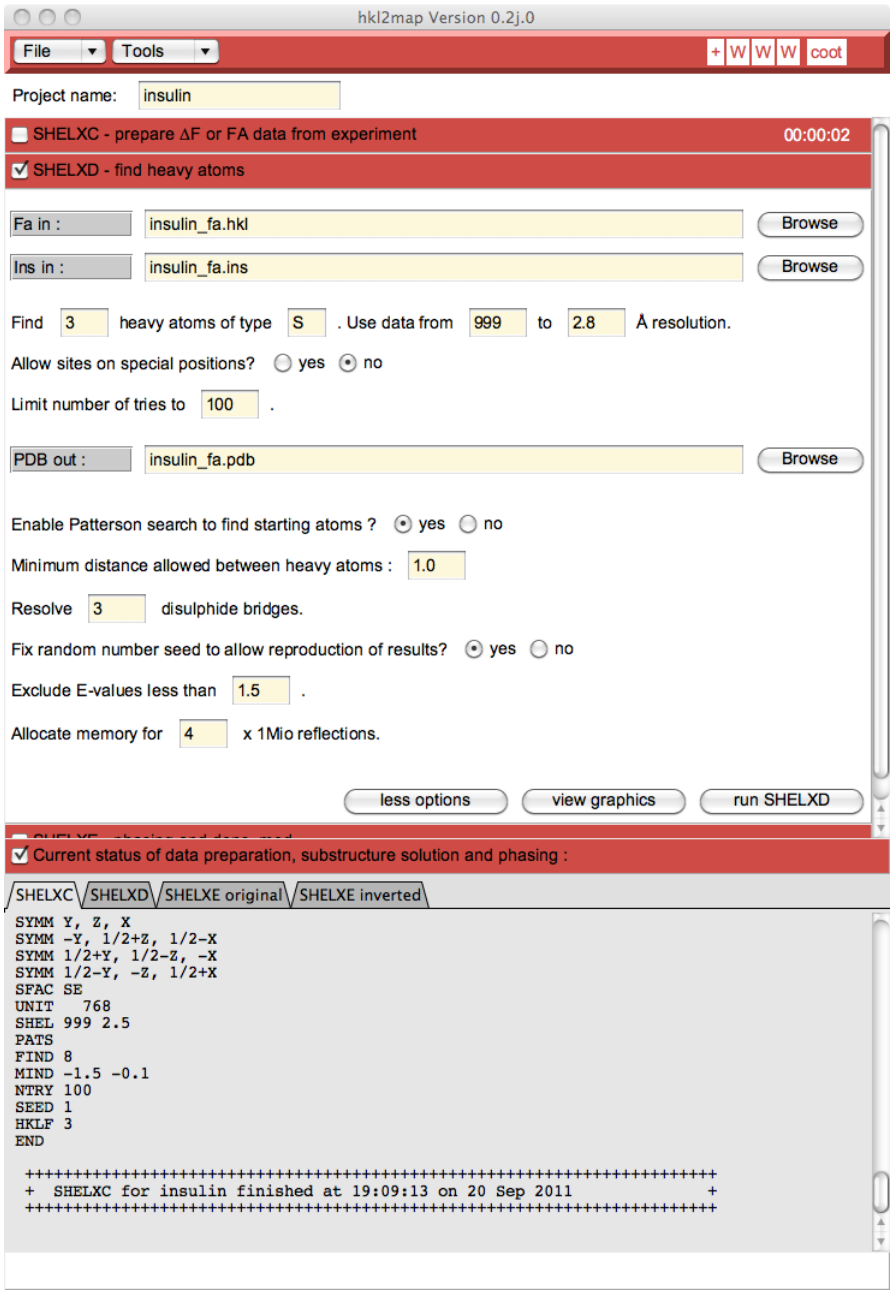

Figure 13: Input parameters for SHELXD to find 6 sulphur atoms in three disulphide bridges using data to 2.8 Å resolution

The best solution has a CC_{all} of 53.3 percent.

5.5 Running SHELXE

Starting SHELXE using the sites determined by SHELXD gives a $CC(E_{obx}, E_{calc})$ of 25.22% indicating that things go well.

After a few cycles of density modification, contrast and Pseudo-free CC clearly indicate that the original substructure has the right hand for the phase calculation.

In fact, after 3 rounds of auto-tracing, 50 out of 51 residues have been traced, resulting in an CC

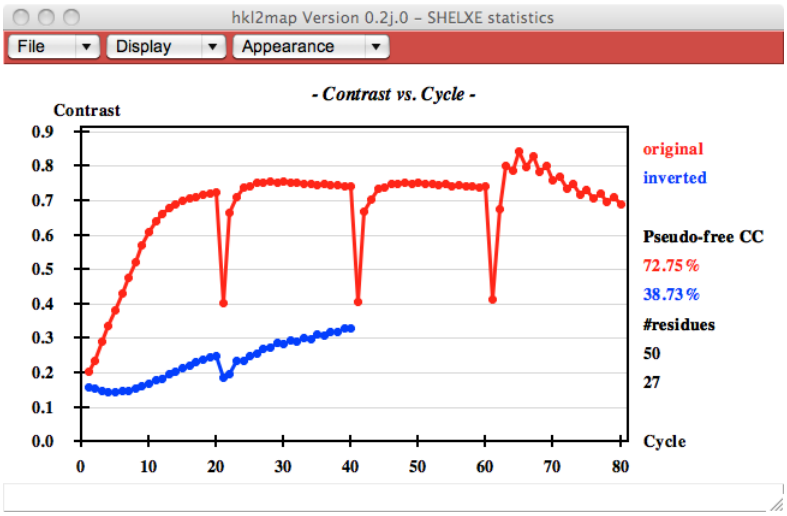

Figure 14: Plot of contrast as a function of cycle number with starting phases calculated based on the original (red) and the inverted (blue) substructure.

for the partical structure against the native data of 53.13%

6 Various

6.1 Tips and Tricks

6.1.1 How can I import data from mtz-format?

Use Time Grünes program `mtz2sca`. This program can be obtained from `http://shelx.uni-ac.gwdg.de/~tg/mtz2x/mtz2sca/download.php` Using `mtz2sca`, you can easily convert `mtz`-files into `sca`-files that can be read by HKL2MAP.

6.2 Known issues

1. 110827. Coot 0.6.1. The .hat-file written by SHELXE-110819 contains two END statements. This is valid SHELXE-syntax. The sites between the END-statements are sites considered useless by SHELXE. Such a .hat file crashes coot. Therefore the current version of HKL2MAP does not automatically load the .hat file into COOT when COOT is started.
2. 110929. Tcl/Tk on MACs. On MACs, the initial directory for file browsing may be a strange place. This is not a bug, but a features of the MAC. Apparently, there is some switch in the system preferences which controls this. I was not able to find that switch.
3. Finishing SHELXE upon request when auto-tracing is activated can take quite a while, because SHELXE will finish the current cycle of auto-tracing and the subsequent density modification before stopping.

7 Acknowledgements

Thanks to Zbyszek Dauter for providing the TE-II-data. The making of this tutorial was in part supported by the EU-funded projects AUTOSTRUCT (QLRI-CT-2000-00398) and BIOXHIT.

References

[1] J. Li, U. Derewenda, Z. Dauter, S. Smith, and Z.S. Derewenda. Crystal structure of the escherichia coli thioesterase ii, a homolog of the human nef binding enzyme. *Nat.Stru.Biol.*, 7:555–559, 2000.